TAXON 23(2/3): 291-300. MAY 1974

PHYLOGENY OF PROCARYOTES

*J. De Ley**

Summary

We first list the main experimental methods used to get an insight into the phylogenetic relationships of the presently existing bacterial genera. Their scope and relative merits are briefly discussed.

As a model system we briefly summarize the application of most of those methods to elucidate the phylogenetic relationships within the genus *Agrobacterium*. The evolutionary development of this genus is described.

Similar experiments with *Zymomonas* are summarized.

Introduction

The phylogeny of procaryotes, in particular of bacteria, remained an enigma until about ten years ago. The phylogeny of the blue-green algae is still an enigma.

In this paper we shall be concerned solely with bacteria. The relative lack of knowledge on the phylogeny of these organisms is largely due to their very small size. Phylogenetic relationships of higher organisms have been unraveled because of comparative morphology, anatomy and embryology, of paleontology, stratigraphy and several other disciplines. In the bacterial world, however, diversity in morphology and anatomy is very limited and the other disciplines, mentioned above, are of no or very limited applicability. Therefore the insight in the phylogeny of the higher organisms is fair to excellent, but for bacteria it remained a near void.

Several schemes of bacterial phylogeny and evolution have been put forward in the past. We quote only the most important ones. Stanier en van Niel (1941) imagined that a coccus was the simplest bacterium from where the others evolved in separate lines. According to Bisset (1962) a spirillum was the primitive starting point for bacterial evolution. Knight (1938) and Lwoff (1943) proposed a picture of bacterial evolution based on a nutritional and physiological background. None of these hypotheses are satisfactory because of the lack of data. From modern molecular-biological experiments on DNA : RNA hybridization (De Ley, De Smedt and others, to be published) it appears that the hypotheses of Stanier and van Niel may contain a nucleus of truth. Recently Margulis (1970, 1972) proposed a phylogenetic tree of bacteria. It contains a number of correlations between bacterial genera, which are not supported by extant physiological, biochemical and molecular-biological data.

* Laboratory of Microbiology and microbial Genetics, Faculty of Sciences, State University, Ledeganckstraat 35, 9000 Gent, Belgium.

Detection of bacterial phylogeny

Experimental data from different approaches are now accumulating, permitting to see phylogenetic relationships within and between several groups of genera, such as e.g. *Escherichia* with *Shigella, Salmonella, Enterobacter* and other *Enterobacteriaceae; Agrobacterium* with *Rhizobium*; *Pseudomonas* with *Xanthomonas,* etc. However, the phylogenetic relationships between most other genera are unknown.

Several methods allow to obtain an insight in phylogenetic relationships amongst bacteria.

1. *The classical or orthodox taxonomy* uses generally at most 25 morphological, physiological, biochemical, serological and ecological criteria for determining a strain. A number of strains have been grouped together in taxa in an almost true picture of their real phylogenetic relationships. In many cases, however, the small number of criteria used (probably only a few per cent of the genetic potential) led to many mistakes and to confusion. Examples of the latter are e.g. the chaotic classification of the socalled genera *Achromobacter, Alcaligenes, Flavobacterium,* etc. There are many examples of mistaken identities, misidentified strains, wrongly labeled taxa, etc. E.g. strains included in *Agrobacterium* as the species *A. gypsophilae* and *A. pseudotsugae*, are in reality *Erwinia herbicola* and a coryneform, respectively.

2. *Adansonian or numerical analysis.* This method uses more strains and more features, involving probably up to one third of the genome. The large number of data are handled by computers and grouped by mathematical clustering methods. The probability is thus much greater that the strains are grouped together according to their real relationship. The groups obtained in this fashion are taxonomically and phylogenetically more significant. This approach is most powerful for closely related organisms. Similarity of phenotypic features does not always mean identity of the genes concerned. E.g. organisms which ferment glucose may do this by at least five different enzymic pathways. Enzymes may have the same function, but be structurally quite different. Therefore numerical analysis probably exagerates the similarity and decreases the phylogenetic distance of remotely related genera.

3. *Comparative metabolism.* Knowledge of the different enzymic pathways of intermediary metabolism quite frequently allows a broad insight in intergeneric relations. This field was first explored by De Ley (1962) where several examples can be found. More recently an interesting comparison of energy metabolism in the bacterial world was presented by Decker et al. (1970). Very many data on a great variety of enzymes and enzymic reactions in many bacteria are dispersed all over the literature. Computer-assisted data storage and comparison would yield much information.

4. *Electropherograms of soluble proteins* (protein band profiles). In this laboratory we worked out a procedure for the comparison of electrophoretic protein band profiles of large numbers of bacteria (Kersters and De Ley, to be published). It consists essentially of a very reproducible technique for making the profiles and a computer-assisted clustering method. We examine thus a kind of protein copy of each genome. Our experience shows that our method has very much resolution. It detects small individual differences between strains and it is able to resolve a species or a small genus into its genetic races. It can detect intergeneric relationships but not quantitatively.

5. *Average DNA base ratios* (average % GC). This subject has been so extensively dealt with in the literature that a few words on it may suffice. The average DNA base ratio is very important as it is one of the very few constants in biology. In itself it does not measure the degree of relationship. When two organisms have an identical or similar % GC, they may be, but they are not necessarily related. Other criteria are then needed. However, when two organisms are 25 or more % GC apart, their genomes are mutationally widely different or even totally unrelated. Several examples have been given by De Ley (1969).

6. *DNA:DNA hybridization.* Similarity in genome DNA base sequences are measured by DNA:DNA or DNA:cRNA hybridizations. The interpretation of the different methods is apparently very simple: the fraction of two genomes which hybridizes is considered to be ancestral and a direct measure of relatedness. In reality, the situation is much more complex. Except in very closely related strains (e.g. two strains of *E. coli*) mutations affected the genomes unevenly and often quite drastically in different sites over their length. Mismatchings in heterologous DNA duplexes are inversely proportional to the amount of DNA duplexed. The per cent of mismatchings determines the thermal stability of the hybrids. Unfortunately, there is a strong tendency among some workers to carry out hybridizations at temperatures above T_m-25° C. In this fashion one artificially selects for a stable fraction of the hybrid, which may be unrepresentative of the degree of relationship; the less-stable DNA hybrids are not detected, but they may be phenotypically at least as important. Quantitative conclusions on homology are only valuable for a group of organisms with a high degree of relatedness. Several bacterial taxa have been clarified in this fashion. When organisms have a homology of less than 70% at T_m - 25° C, the experimental degrees of relatedness are too low. This applies to many inter-species, and intergeneric DNA hybridizations which have been reported with the stringent method. For the moment the true genetic similarity between most bacterial genera cannot yet be determined with certainty by DNA:DNA hybridization. Molecular biological aspects of bacterial phylogeny have been discussed by De Ley (1968).

7. *Sequencing of small proteins and RNA molecules.* A number of bacterial cytochromes c (Ambler, personal communication) ferredoxins, and other small molecules have been sequenced. This field is very promising to provide insight into intergeneric evolution. Fragments of the relationships which have been resolved, have been summarized by Dayhof (1972).

Phylogenetic relationships within the genus Agrobacterium

We decided to determine experimentally the phylogenetic relationships of strains within one group of bacteria as a model system. We needed a rather small genus for this study. *Agrobacterium* was selected for several reasons. Firstly, the number of strains known, about 300, is not too large and experimentally manageable. We have most of them in our collection. Secondly, they have been isolated all over the world. Thirdly, this genus seems to be identified phenotypically rather easily. Fourthly, the number of species is rather small: it contains the crown gall bacteria *A. tumefaciens,* the cane gall bacteria *A. rubi,* the hairy root disease bacteria *A. rhizogenes,* and nonpathogenic strains called *A. radiobacter.* About eleven other species such as *A. gypsophilae, A. stellulatum, A. pseudotsugae*, etc. are taxonomic mistakes, belonging in other genera, and will not be considered here. Fifthly, many strains cause the formation of plant tumors. The real molecular cause of this disease is not yet known. A better knowledge of these bacteria is required for the elucidation of the pathogenic mechanism. We used a variety of methods.

1. *DNA:DNA hybridization.* It was our purpose to determine firstly quantitative similarities between genomes. Preliminary research (Heberlein, De Ley, Tytgat, 1967) showed that the genus is genetically heterogeneous. To establish the degree of DNA homology amongst e.g. 200 strains, theoretically 20.100 hybridizations are required. Fortunately it turned out we could do with much less, but still a few thousands, including the controls, were required. We first developed a new method, based on initial renaturation rate measurements,

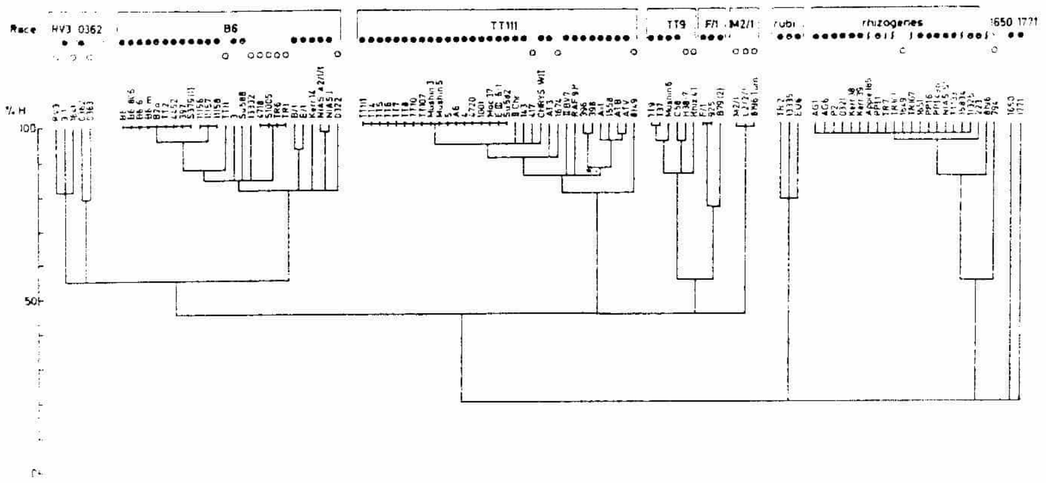

Fig. 1. Relatedness in the genus *Agrobacterium*, determined with DNA: DNA hybridization.
This tree was determined with the initial renaturation rate method (De Ley et al., 1970). Many data were confirmed with the DNA membrane filter method. The existence of the eleven genetic races is obvious. Symbols: • tumorigenic strain, ° doubtfully tumorigenis strain, s excessive root-forming strain. The figure is extended and modified from De Ley (1972).

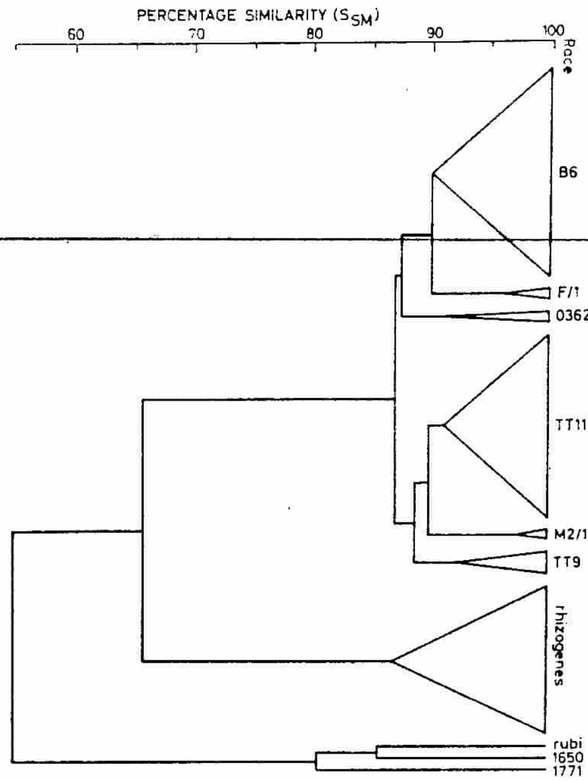

Fig. 2. Numerical analysis of 120 phenotypic features in 70 *Agrobacterium* strains. Unweighted pair group method. The full data are given by Kersters, De Ley, Sneath and Sackin in J. gen. Microbiol. 1973.

which required no radioactive DNA and took only 0.5 hr instead of the usual 16 hr (De Ley, Cattoir and Reynaerts, 1970). If classical taxonomy is correct, one would expect four groups of hybrids, each corresponding to a species. Reality is quite different (fig. 1). The genus consists of 11 genetic races. Three are large and contain each about 70 strains. They are called here the B6, the TT111 and the rhizogenes race. The remaining eight races contain between 1 and 10 strains. Within each race the strains are at least 80% DNA-homologous. The B6 race, the TT111 race and five small races hybridize at ca 50% homology. We call it the tumefaciens-radiobacter cluster. This large aggregate, the rhizogenes race, the rubi race and two individual strains 1650 and 1771 hybridize with about 15% DNA homology. Hybridizations were carried out at T_m - 25°. Other temperatures are being explored. It is quite clear that many mutations happened within this genus. We shall see below that these genetic races are real, but that the phylogenetic distances are exagerated. One expects that strains, which are 50% or 20% DNA-homologous should show considerable phenotypic differences.

2. *Numerical analysis of phenotypic features* (Kersters, De Ley, Sneath and Sacking, J. gen. Microbiol. 1973). We used a random selection of 70 strains. From the *Agrobacterium* literature we chose over a hundred different phenotypic features, such as growth on amino acids, organic acids, aromatic compounds, growth and acid production on a series of carbohydrates, 3-ketolactose test, litmus milk test, nitrate and nitrite reduction, phytopathogenicity, etc. The presence or absence of each feature was determined in each strain. The results are represented in Fig. 2. Organisms within each genetic race show no significant phenotypic differences. Genetic races at the 50% homology level differ by at most 5 features. The rhizogenes race differs by some 20 features from the tumefaciens-radiobacter cluster. The three other races at 15% homology differ more phenotypically. One clearly sees, that even in a well-defined and homogeneous genus, DNA hybridization in stringent conditions diminishes phylogenetic relationships. The % similarity levels, calculated by numerical analysis are > 93, > 85 and > 65%, the homology levels are > 80, 50 and 15%. This quantitative paradox can be explained if mutations are located mainly outside the active center of the enzymes.

3. *Electrophoretic protein profiles* (Kersters and De Ley, to be published). If the above hypothesis is correct, one should expect great differences in the electrophoretic protein profiles between the different genetic races. The experiments confirm this hypothesis completely. We find that strains of each genetic race display similar protein patterns and that each race has its own typical pattern. One example has been given by De Ley (1972). The phenotypic explanation has thus been given. The electropherograms can replace DNA hybridization techniques in this genus; they are extremely well suited as a simple and quick identification method.

4. *Racial differences between cytochrome c* (Van Beeumen, Van den Brande and De Ley, to be published). The previous conclusions call for a direct

Table 1. Some properties of pure cytochrome c from strains of the three main genetic races in *Agrobacterium* (Van Beeumen, Van den Brande en De Ley, to be published).

Genetic race	B6		TT111	rhizogenes
strain	B2a	B6	II Chrys	Apple 185
$\lambda\alpha$	556	556	556	556
Mol. weight in SDS	11.000	11.000	11.000	11.000
Isoelectric point	5.67	5.67	6.23	5.27
el. phor. mobility	0.57	0.57	0.40	0.73

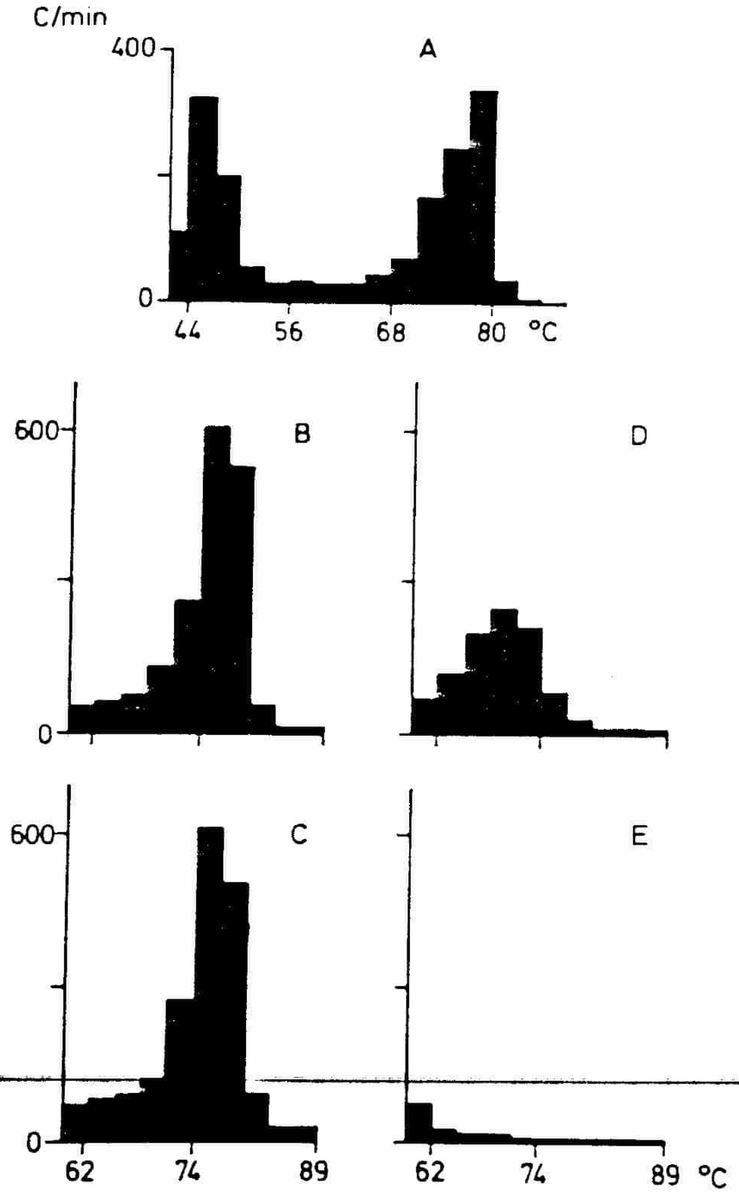

Fig. 3. Thermal stability of DNA: DNA hybrids in *Agrobacterium.*

Fig. A. The presence of a labile and a stable hybrid fraction in a heterologous hybridization at T_m–42° C.

Fig. B.-E. different types of stable hybrids formed at T_m–25° C. Fig. B. a control homologous renaturation for comparison. Fig. C. the stability of the hybrid formed between two DNA's of strains within the same race. Fig. D. idem between two DNA's of strains from races with 50% homology. Fig. E. idem between two DNA's of strains from races with ca 15% homology. The full data by De Ley et al. are given in J. gen. Microbiol. 1973.

experimental confirmation. The obvious choice is the comparison of amino acid sequences of cytochrome c from each of the main races. We prepared cytochrome c 556 from the races B6, TT111 and rhizogenes in pure state. Table 1 summarizes some of their features. The differences in isoelectric point and in electrophoretic mobility strongly suggests that there must be differences in amino acid composition. These experiments are now being extended.

5. *DNA base composition and genome size.* All strains have a % GC within the range 58 to 62. There are few noticeable differences between the genetic races.

The molecular weight of the genome DNA of many strains was determined with a new method developed in our laboratory and based on initial renaturation rates (Gillis, De Ley and De Cleene, 1970). The genome DNA of strains from each of the three main races is the same within the experimental error. The average molecular weight is 3.4 x 10^9, corresponding to about 5.6 million base pairs. The *Agrobacterium* genome is about 40% longer than the *E. coli* genome. Genomes of some of the smaller races have sometimes a slightly (up to 15%) different molecular weight. In summary one can say that the size and base composition of the *Agrobacterium* genome has not changed noticeably during evolution.

6. *Thermal stability of DNA:DNA hybrids* (De Ley, Tytgat, De Smedt and Michiels, J. gen. Microbiol. 1973). Up to 80% of the genome length has been sufficiently changed by mutation to prevent hybridization. One can wonder whether the common parts which constitute the hybrids, have been affected by mutation, and if so, how much. This can be determined from the thermal stability curve of the hybrid (Fig. 3). Depending on the temperature used two types of hybrids are formed (Fig. 3A). One is a labile one of unknown nature which will not be considered here. The other type is more or less stable, depending on its origin. Within each of the main *Agrobacterium* races thermal stability of the DNA hybrid was about the same, within 2° C, as for the homoduplex (Fig. 3B and C). Between two races of 50% homology the duplexes were about 6° C less stable (Fig. 3D). Between races of 15% homology the stability was at least 13° C lower (Fig. 3E). Obviously, the less two races are evolutionary related, the more

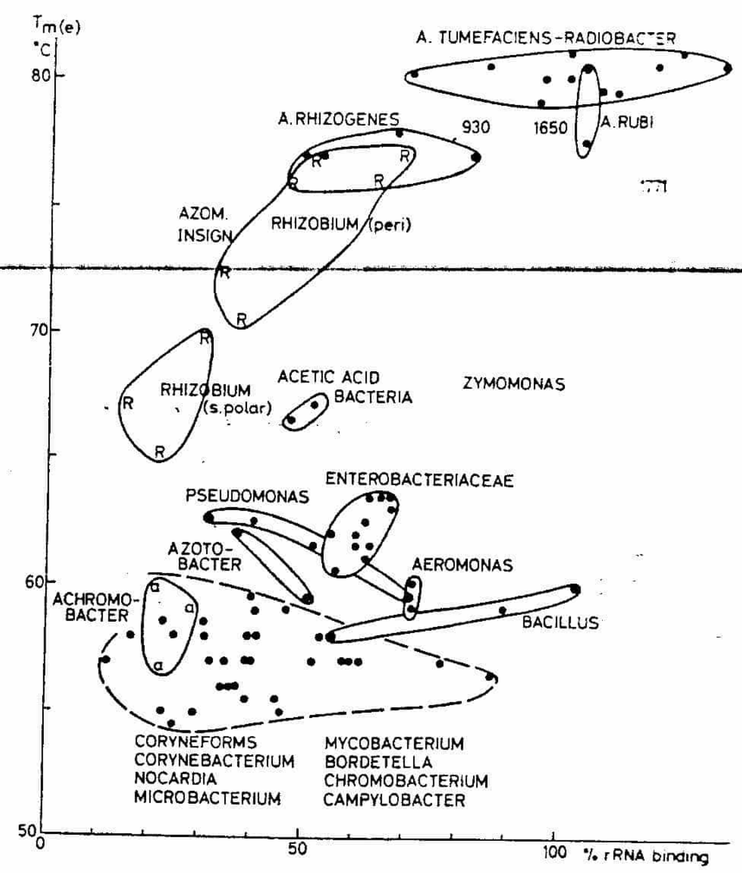

Fig. 4. The amount and the thermal stability of heterologous hybrids between ^{14}C-rRNA from *Agrobacterium tumefaciens* TT111 with DNA from many other bacteria.

mutations occurred within the common part. DNA relatedness between the B6, TT111 and some of the small races show that single-stranded stretches, totalling about 2.8 x 10^6 bases are still able to hybridize. About 10^5 of these bases would be mismatched. The duplexing stretches between the rhizogenes and the other races total about 8 x 10^5 bases. The mismatching would be over 50.000 bases.

7. Intergeneric relationships and the evolutionary origin of Agrobacterium. Previous results (Heberlein et al., 1967) showed that there is a definite DNA relationship between *Agrobacterium* and *Rhizobium*. The relationship with other genera is more remote and hard to detect by DNA:DNA hybridization. In the *Enterobacteriaceae* and a few other organisms the base sequences in the rRNA cistrons are conserved, i.e. they have mutated less than the rest of the genome (e.g. Moore and McCarthy, 1967). It seemed to us that this is an approach to establish remote relationships between bacterial genera. A great number of intergeneric DNA:rRNA hybridizations substantiated this view (De Smedt and De Ley, to be published). We report here only on hybridizations with ^{14}C-rRNA from *A. tumefaciens* TT111 as an example (Fig. 4). One can distinguish, but less clearly, the same subdivision in *Agrobacterium* as with the previous methods. The rRNA cistrons from the peritrichously flagellated fast growing *Rhizobium* strains are almost identical. The rRNA cistrons of the subpolarly flagellated, slow growing rhizobia, the acetic acid bacteria and *Zymomonas* are already clearly different. Other genera such as *Pseudomonas*, *Bacillus*, the *Enterobacteriaceae*, many gram positives and other mentioned in Fig. 4, are only remotely related.

8. The evolution of Agrobacterium. We have now enough data to visualize the origin of *Agrobacterium* and the evolution within this genus. After the rise of the dicotyledonous plants some 200 million years ago, the ancestors of the present rhizobia and agrobacteria probably started to develop in two directions.

One group of bacteria acquired the property to grow in many plants and to produce tumors by an as yet unknown mechanism. They were the ancestors of the presently living *Agrobacterium*.

The second group started to live symbiontically with the leguminous plants and were probably responsible for their present great importance. These bacteria formed at least two branches, the ancestors of the presently living subpolarly and peritrichously flagellated rhizobia. Alternatively it is possible that the present agrobacteria arose from existing ancestral rhizobia or vice versa. But it is certain that they are all twigs of the same branch.

Let us consider here only the ancestral agrobacteria. The first steps in the evolution within this group probably was nearly the same as we see now within each genetic race: a limited number of mutations, mainly outside the active center of enzymes, perhaps mainly in the third base of the triplets or permitting substitution of similar amino acids in a few proteins. Their phenotype was almost identical. For some unknown reason (geographical, ecological?) this population was dispersed into at least five groups, the ancestors of the present tumefaciens-radiobacter cluster, the rhizogenes, rubi, 1650 and 1771 races. Under different conditions they evolved away from each other, probably very similar to the differences between the present B6 and TT111 races: more mutations modifying the structural features of many proteins, mutations in active centers of a few enzymes. Their phenotype was first slightly different. More and more mutations affected active centers until the groups reached their present diversity: ca 15% DNA homology, considerable mutations in the common DNA parts and at least 20% difference in their phenotypic features. The races rubi, 1650 and 1771 remained extremely small in number. The rhizogenes race spread out over the entire world, mainly in warm or arid climates; some strains developed an unknown system to induce excessive root growth. This

race is either genetically very stable or its geographical spreading occurred rather recently, because almost all strains are genotypically and phenotypically very similar. The tumefaciens-radiobacter cluster started a further evolutionary diversification all on its own. It evolved into seven races, all with 50% DNA homology, moderate mutations in the common parts, considerable mutations outside the active center changing the properties of proteins, and less than 5% differences in their phenotype. Two races spread out. The B6 race apparently prevails around the Pacific Ocean, as it was isolated so far only in the U.S.A., Japan, New Zealand and Australia. The TT111 race spread out around the Atlantic Ocean: it was isolated in the U.S.A., Europe, South Africa. The five other races remained small in number of strains. Most of them started to diversify already slightly, as testified by their 80-100% DNA homology.

Fig. 1 also shows that there is no correlation between the species and the genetic races. Strains named *A. tumefaciens* occur in all races,. strains named *A. rubi* form one race with strains named *A. tumefaciens*, *A. radiobacter* occurs in seven races, strains named *A. rhizogenes* and *A. tumefaciens* occur at very high DNA homology in the same race. It is obvious that the present species concept and species differentiation in *Agrobacterium* is not justified. A nomenclatorial revision will have to be carried out. We can now also assess the value of the four current Latin species names: they refer only to the state of phytopathogenicity, to the presence or absence of a disease-provoking property; they have only the value of formae speciales, pathotypes; they do not represent a true biological unit.

Phylogenetic relationships within the genus Zymomonas

Not all bacterial taxa diversified in the same fashion as *Agrobacterium*. Some populations diversified phenotypically much more profoundly, as in the family of the *Enterobacteriaceae*. Others remained more homogeneous as we shall see in the next example of *Zymomonas*.

These gram-negative anaerobic rods ferment glucose or fructose through a modification of the Entner-Doudoroff mechanism to ethanol, CO_2 and a little lactic acid. They have been isolated from fermenting plant material and in breweries. We have many of the existing strains in our collection: they are from fermenting agave juice (pulque, Mexico), palm wines (Zaïre) and breweries (England). We studied the same parameters as for *Agrobacterium* (Swings and De Ley, to be published). We shall report our detailed results elsewhere. *Zymomonas* is a very tight group with over 80% DNA homology (except one strain), almost identical phenotypic features and protein profiles, % GC and genome size. These bacteria, because of their strict requirements (presence of glucose or fructose, growth factors, amino acids and anaerobiosis) occur only in approximately the same selective conditions. They hardly diversified except in minor features, whether they occur in Zaïre, in Mexico or in European breweries. So far we have not yet found generic relatives. Apparently this genus evolved and remained all by itself in the large group of gram-negative bacteria.

Acknowledgement

We are indebted to the Fonds voor Kollektief Fundamenteel Onderzoek (Belgium) for supporting our research with several grants.

References

Bisset, K. A. 1962 – The phylogenetic concept in bacterial taxonomy. 12th Symp. Soc. Gen. Microb. "Bacterial Classification" 361-373. Cambridge University Press.

Dayhoff, M. O. 1972 – Atlas of protein sequence and structure. NBFR, Washington.

Decker, K., Jungermann, K. & R. K. Thauer 1970 – Wege der Energiegewinnung in Anaerobiern. Angew. Chem. 82: 153-173.

De Ley, J. 1962 – Comparative biochemistry and enzymology in bacterial classification. 12th Symp. Soc. Gen. Microb. "Bacterial Classification" 164-195. Cambridge University Press.

De Ley, J. 1968 – Molecular Biology and bacterial Phylogeny. Evolutionary Biology. Vol. 2. Ed. Th. Dobzhansky, M. K. Hecht and Wm. C. Steere. 103-156. Appleton-Century-Crofts. New York.

De Ley, J. 1969 – Compositional nucleotide distribution and the theoretical prediction of homology in bacterial DNA. J. Theor. Biol. 22: 89-116.

De Ley, J. 1972 – *Agrobacterium:* intrageneric relationships and evolution. Proc. 3d Int. Conf. Plant Pathog. Bacteria. April 1971. Wageningen, the Netherlands, Pudoc.

De Ley, J., Cattoir, H. & A. Reynaerts 1970 – The quantitative measurement of DNA hybridization from renaturation rates. Eur. J. Biochem. 12: 133-142.

De Ley, J., Tytgat, R., De Smedt, J. & M. Michiels 1973 – Thermal stability of DNA : DNA hybrids within the genus *Agrobacterium.* J. Gen. Microbiol. 78: 241-252.

Gillis, M., De Ley, J. & M. De Cleene 1970 – The determination of molecular weight of bacterial genome DNA from renaturation rates. Eur. J. Biochem. 12: 143–153.

Heberlein, G., De Ley, J. & R. Tytgat 1967 – Deoxyribonucleic acid homology and taxonomy of *Agrobacterium, Rhizobium* and *Chromobacterium.* J. Bacteriol. 94: 116-124.

Kersters, K., De Ley, J., Sneath, P. H. A. & M. Sackin 1973 – Numerical Taxonomic Analysis of *Agrobacterium.* J. Gen. Microbiol. 78: 227-239.

Knight, B. C. 1938 – Bacterial Nutrition. H. M. Stationery Office, London.

Lwoff, A. 1943 – L'évolution physiologique. Hermann et Cie, Paris.

Margulis, L. 1970 – Origin of Eukaryotic Cells. Yale University Press, New Haven.

Margulis, L. 1972 – Early cellular evolution. Exobiology, Ed. C. Ponnamperuma. North Holland Public. Co. 342-368. Amsterdam.

Moore, R. L. & B. McCarthy 1967 – Comparative study of ribosomal ribonucleic acid cistrons in Enterobacteria and Myxobacteria. J. Bacteriol. 94: 1066-1074.

Stanier, R. Y. & C. B. van Niel 1941 – The main outlines of bacterial classification. J. Bacteriol. 42: 437-466.